

# *A spatial-and-temporal-based method for rapid particle concentration estimations in an urban environment*

Article

Accepted Version

Creative Commons: Attribution-Noncommercial-No Derivative Works 4.0

Xiong, J., Yao, R., Wang, W., Yu, W. and Li, B. (2020) A spatial-and-temporal-based method for rapid particle concentration estimations in an urban environment. *Journal of Cleaner Production*, 256. 120331. ISSN 0959-6526 doi: <https://doi.org/10.1016/j.jclepro.2020.120331> Available at <https://centaur.reading.ac.uk/88951/>

It is advisable to refer to the publisher's version if you intend to cite from the work. See [Guidance on citing](#).

To link to this article DOI: <http://dx.doi.org/10.1016/j.jclepro.2020.120331>

Publisher: Elsevier

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the [End User Agreement](#).

[www.reading.ac.uk/centaur](http://www.reading.ac.uk/centaur)

**CentAUR**

Central Archive at the University of Reading

Reading's research outputs online

# A spatial-and-temporal-based method for rapid particle concentration estimations in an urban environment

Jie Xiong<sup>1,2</sup>, Runming Yao<sup>1,3\*</sup>, Wenbo Wang<sup>3</sup>, Wei Yu<sup>1,2</sup>, Baizhan Li<sup>1,2\*</sup>

<sup>1</sup> Joint International Research Laboratory of Green Buildings and Built Environments (Ministry of Education),  
Chongqing University, Chongqing 400045, China

<sup>2</sup> National Centre for International Research of Low-carbon and Green Buildings (Ministry of Science and Technology),  
Chongqing University, Chongqing 400045, China

<sup>3</sup> School of the Built Environment, University of Reading, Reading RG6 6DF, UK

\* Corresponding author: [r.yao@reading.ac.uk](mailto:r.yao@reading.ac.uk); [baizhanli@cqu.edu.cn](mailto:baizhanli@cqu.edu.cn)

## Abstract

The increasing construction of buildings and infrastructure in cities heavily influences pollutant dispersion and causes a spread of increased particle concentrations. Real-time data and information on local pollution levels are highly desired by residents, urban planners and policy-makers. Such information is scarce due to the high cost of real-time measurement. To fill the gap, the aim of this research is to develop a model that can rapidly estimate particulate pollution based on a data-driven artificial neural network modelling approach. The key influential factors such as background pollution level, weather conditions, urban morphology and local pollution sources are embedded in the model in association with local emission sources of pollution relating to construction activities and traffic flows. The data for urban spatial-variables (building and road) and traffic information is processed with the aid of the Geographic Information System using self-developed Python scripts. The geographic dataset containing the required information for each grid is integrated with the artificial neural network model to perform forecasting of particle

concentrations. The model has been verified with measurements from a case study with 20 sample locations in Chongqing, China, showing that the average relative error of particle concentration estimation compared to measurement is 17.56% for PM<sub>10</sub> and 16.04% for PM<sub>2.5</sub>. A map of a time-specific spatial interpolation of particle concentrations which visualises real-time pollution is consequently produced based on the method. The method can be used as a tool for real-time air quality forecasting with suitable adaptations for any other dense urban area with minimum information from local observation stations.

**Keywords:** Particulate matter; Artificial Neural Network (ANN); Urban morphology; Traffic emissions; Geographic Information System (GIS); Spatial interpolation

## Acronyms

<i>ANN</i>	Artificial Neural Network
<i>API</i>	Air Pollution Index
<i>CFD</i>	Computational Fluid Dynamics
<i>GIS</i>	Geographic Information System
<i>MLR</i>	Multiple Linear Regression
<i>PCA</i>	Principal Component Analysis
<i>PM</i>	Particulate matter, also Particle
<i>SLR</i>	Simple Linear Regression
<i>WHO</i>	World Health Organization

## Nomenclature

$a_j^l$	The $j^{\text{th}}$ neuron in the $l^{\text{th}}$ layer
$A_{cs}$	Area of the construction site (m <sup>2</sup> )
$A_i$	Coverage area of the building $i$ (m <sup>2</sup> )

$b_j^l$	Bias of the $j^{\text{th}}$ neuron in the $l^{\text{th}}$ layer
$Bias$	Average bias
$BCR$	Building coverage ratio
$BH$	Coverage-area-weighted average building height (m)
$CS_t$	Average congestion status in a land lot (0, 1.0~4.0)
$D_{cs}$	Distance of nearest construction site (m)
$D_i$	Distance to the nearest main road (m)
$f(*)$	Activation function
$hh$	Hour sequence in a day
$h_i$	Height of the building $i$ (m)
$L_i$	Length of the road $i$ (m)
$LC_t$	Lane-count of the nearest main road
$m$	Total number of roads in the target area
$\bar{M}$	Average of measured values
$M_i$	The $i^{\text{th}}$ measured value
$n$	Total number of building in the target area
$N_i$	Number of lanes for the road $i$
$\bar{P}$	Average of predicted values
$P_i$	The $i^{\text{th}}$ predicted value
$r$	Pearson correlation coefficient
$RF$	Precipitation (mm)
$RH$	Relative humidity (%)
$RMSE$	Root mean square error
$S$	Total land area of the target (m <sup>2</sup> )
$SL_t$	Speed limit of the nearest main road (km.h <sup>-1</sup> )
$SLRL$	Single-lane road length per unit area (km.km <sup>-2</sup> )
$Temp$	Temperature (°C)

$w_{jk}^l$	Weight for the connection from the $k^{\text{th}}$ neuron in the $(l-1)^{\text{th}}$ layer to the $j^{\text{th}}$ neuron in the $l^{\text{th}}$ layer
$W$	Day sequence in a Week
$WS$	Wind speed ( $\text{m.s}^{-1}$ )

## 1 Introduction

Cities and towns accommodate people to live, study, work and entertain. The scale and speed of global urbanisation have drawn research attention towards the issue of air pollution. The outdoor atmospheric environment mainly contains particulate matter (PM), ozone ( $\text{O}_3$ ), nitrogen oxides ( $\text{NO}_x$ ), sulphur dioxide ( $\text{SO}_2$ ) and other pollutants (World Health Organization, 2006). Airborne particles, existing across a wide range of size with diameter from  $>100\mu\text{m}$  to  $<0.1\mu\text{m}$ , can be categorized in terms of aerodynamic diameter, which determines where the particles can penetrate human organs.  $\text{PM}_{10}$  with an aerodynamic diameter that is generally  $10\mu\text{m}$  and smaller possibly enters the lungs;  $\text{PM}_{2.5}$  with an aerodynamic diameter that is less than  $2.5\mu\text{m}$  possibly enters the bloodstream (United States Environmental Protection Agency, 2018). Some of these particles are emitted directly from sources, such as construction sites, unpaved roads, or fires, but some particles form in the atmosphere resulting from some complex chemical reactions. Thus, PM has a complicated composition made up of hundreds of substances categorised as inorganic particles, organic particles and living particles, which makes them of greater health significance than any other air pollutants. The consequences arising from the entry of PM into the human body are determined by the composition of, and exposure to, the PM. Overall, recent epidemiological studies have confirmed that inhaling PM can cause asthma (Kim *et al.*, 2013; Künzli *et al.*, 2000), lung cancer (Pope III *et al.*, 2002), gastric cancer (Weinmayr *et al.*, 2018), cardiovascular diseases (Künzli *et al.*, 2000; Nayebar *et al.*, 2019; Pope III *et al.*, 2002), respiratory diseases (Guilbert *et al.*, 2019; Künzli *et al.*, 2000), preterm birth (Li *et al.*, 2017), birth defects (Z. Li *et al.*, 2019), premature death (Künzli *et al.*, 2000; Lelieveld *et al.*, 2015) and similar health effects.

In recent years, there is a growing need by the public for informed knowledge on outdoor particle pollution and its impact on human health. In the built environment, natural ventilation, as

one of the powerful passive measures for low energy building design, encountered many challenges due to the outdoor pollution (Costanzo *et al.*, 2019; Tong *et al.*, 2016; Yao *et al.*, 2018). The quantification of pollution concentrations is essential for risk assessment of some environmental-related diseases (Künzli *et al.*, 2000). However, there is a lack of practical methods of providing spatial- and temporal-based quantitative particle concentrations using the limited information available from public sources.

### 1.1 Literature review of prediction methods

There are two main approaches to acquire particle concentration levels: on-site measurements and modelling predictions. The on-site measurement method is highly accurate as it directly reflects the true value of the sampling point when ignoring any system errors. Many cities in the world have official pollution observation stations providing overall ambient air quality information (China National Environmental Monitoring Centre; Department of Environment, Food & Rural Affairs). They provide reference values for a region, known as the *background* pollution level in this article. However, the cost of on-site measurement, including sensors, maintenance and labour, is very high (Mihăiță *et al.*, 2019), which makes it impractical to take measurement everywhere. Additionally, it is unable to measure in an occasion when it does not occur. The modelling prediction method has made up for those defects, and it is further classified into two types: 1) high-dimension, process-driven, physical models and 2) low-dimension, data-driven, statistical models.

The physics-based model, normally the numerical model of particle dispersion, simulates the dispersion process based on basic computational fluid dynamics (CFD) theory and the mass transfer mechanism; it demands sufficient knowledge of microclimate conditions, particle emission sources and the explicit description of physical deposition and chemical transformation processes (Lateb *et al.*, 2016; Li *et al.*, 2006). This method is mostly used to analyse the pollutant dispersion around buildings from certain known sources (Ai and Mak, 2013; Short *et al.*, 2018). Several studies that have used CFD techniques to predict pollutant concentration have focused on the street canyon (Blocken *et al.*, 2012; Tominaga and Stathopoulos, 2011; Vicente *et al.*, 2018). The direct dust

emissions from vehicles provide the main source of data in the model (B. Li *et al.*, 2019) along with consideration of the by-products from chemical reactions (Kim *et al.*, 2019). Assumptions of boundary conditions and estimations of some parameters, like the deposition rate or transformation rate, are crucial and can cause rather large biases for different schemes (Stern *et al.*, 2008). The computation time is usually significant depending on the specific model and hardware capacity (Salim *et al.*, 2011), making it unlikely to provide full time-series data.

In recent years, low-dimensional, data-driven modelling is being favoured due to its highly efficient simulation based on the established relationships between variables and responses, while ignoring the limited knowledge of the processes involved. The multiple linear regression (MLR) and the artificial neural network (ANN) are mainstream approaches to handle the pollutant concentration estimation. MLR is a simple and straightforward way to explain the relationship between one continuous dependent variable and some independent variables. It is very important to recognise that some variables lack multicollinearity (Shieh and Fouladi, 2003). To be more concise, it comes to the simple linear regression (SLR), where the independent variable should be a synthetic and representative index. Zhou *et al.* (2018) applied the SLR to evaluate the relationship between the Air Pollution Index (API) and 7 indices related to urban size, urban shape irregularity and urban fragmentation. He *et al.* (2015) used the vehicle count, traffic-light period, wind speed, temperature and relative humidity to predict particle concentrations at an urban intersection, and combined the MLP model and principal component analysis (PCA) to improve the predictive accuracy of the time-series PM concentration.

For non-linear features, the ANN model inspired by the biological neural network that constitutes animal brains shows better performance (Haykin, 2009). Özdemir *et al.* (2014) and Chaloulakou *et al.* (2003) investigated the relationships between PM<sub>10</sub> levels and meteorological factors (including surface temperature, relative humidity, and wind speed and direction) by comparing ANN models and MLR models, whose results demonstrate that ANNs can provide adequate solutions to demands for predictions of particulate pollution.

Some studies used historical measurement data to predict current and even future data. For example, Ishak *et al.* (2016) and Saeed *et al.* (2017) used historical observations by two popular



statistical learning methods: the support vector machine and the random forest. Perez and Reyes (2001) confirmed that the information extracted from the PM<sub>2.5</sub> time series may be used to implement a neural network architecture in order to make predictions of this quantity several hours into the future whilst others recognised some influencing factors, using the data at that time to make predictions. The main step for this strategy is to determine the predictors (known as features in computer science) and prepare a representative training dataset, in order to provide sufficient information for the networks (Deligiorgi and Philippopoulos, 2011; Shieh and Fouladi, 2003). Most studies considered the relation between particle concentration and meteorological parameters (Chaloulakou *et al.*, 2003; Özdemir and Taner, 2014). He and Liu (2012) added the traffic volume factor into a statistical distribution model - the goodness-of-fit test - to find the lognormal distribution of PM concentration due to the change of traffic volume between morning and afternoon. Honarvar and Sami (2019) further considered the road network structure data to predict the PM concentration based on a transfer learning perspective in which a neural network and regression was leveraged as the core of the prediction. The urban morphology also influences the dispersion of particles, Gennaro *et al.* (2013) developed the ANN model to forecast PM<sub>10</sub> daily concentrations in two contrasting environments: a regional background site and an urban background site, with local meteorological data and information about the origin of air masses being used as inputs. The model performance showed better results for the regional background site than for the urban site because of the unexpected local sources in the urban background site that sometimes occurred. Reasonable inclusion of closely related factors can increase the accuracy of the model's predictions. So far, a holistic method to quantify particle concentrations in a dense urban area simultaneously considering the overall urban pollution level, meteorological conditions, urban morphology and local pollution sources is lacking.

## **1.2 Aim and scope**

The aim of this research is to develop a spatial-and-temporal-dependent model that can quickly estimate PM concentrations at any time and location within an urban area using limited observed

148 data. The ANN model will be applied for its ability to simulate nonlinear functions, to incorporate  
149 various heterogeneous variables and its speed of implementation. Overall pollution level,  
150 meteorological conditions, urban morphology and local pollution sources are all considered within  
151 the model for their close relationship to the particle concentration. All the data for the prediction can  
152 be accessed from a ready-made, real-time, data platform released for the public after digital  
153 processing. The beneficiaries will be threefold: 1) residents can take necessary protective actions; 2)  
154 policy-makers and planners use policy instruments to control pollution; and 3) building end-users  
155 and facilities managers can effectively operate ventilation systems.

## 157 **2 Methods**

158 The ANN method is attempted in the development of an urban air pollution distribution model  
159 that provides particle concentration as the targeted output. The major process of this method is to  
160 identify the predictors that significantly influence the outputs. The research framework is described  
161 in Figure 2. As shown in the figure, there are four steps: a) data collection of predictors (Step 1), b)  
162 field measurements of particles (Step 2), c) the modelling process and verification (Step 3) and d)  
163 application for estimations (Step 4). Finally, a case study area located in Chongqing, China, is  
164 selected to demonstrate the process involved in the development of the method.

165

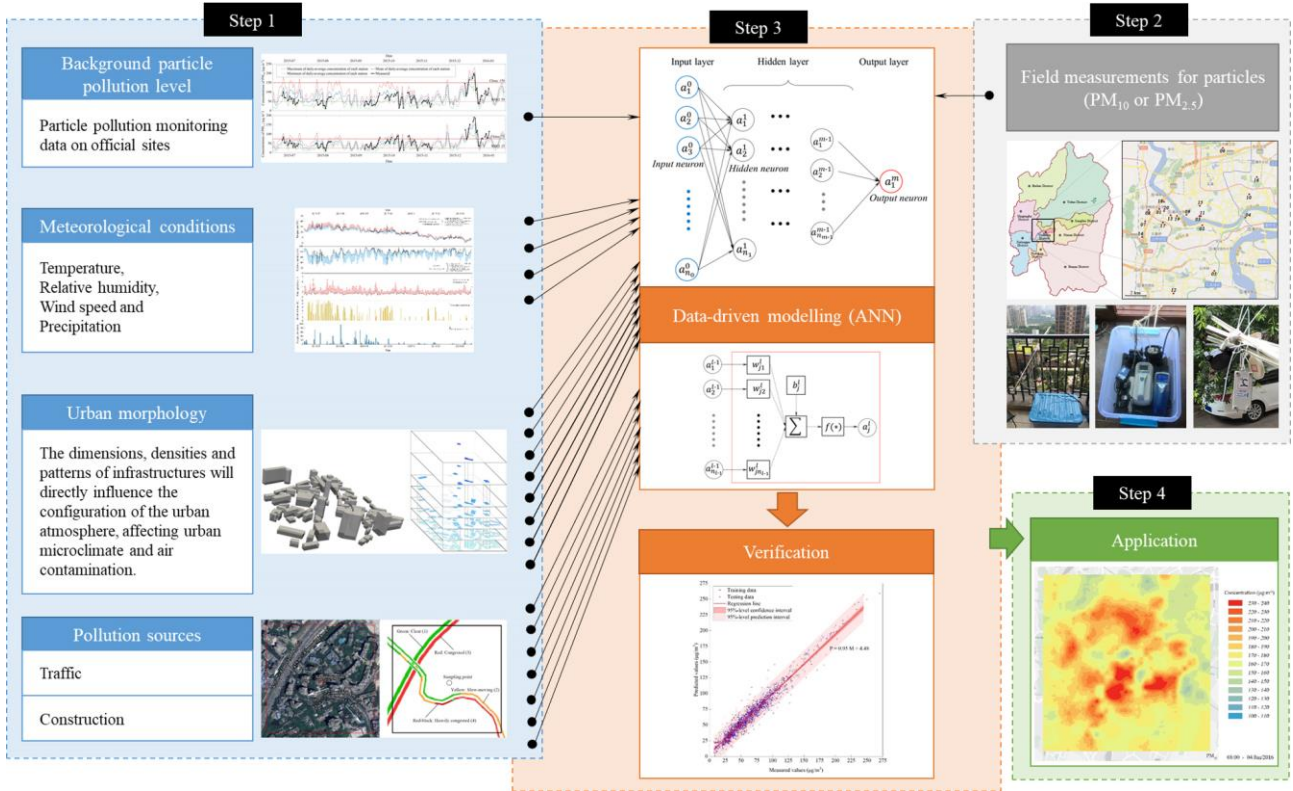


Figure 1: The framework of this research.

## 2.1 Predictors (Step 1)

Determining the predictors and preparing a representative training dataset is key to successfully training an ANN model that can run accurately. Through the analysis of the dispersion process of PM in the UCL (Oke *et al.*, 2017), some main factors affecting the local particle concentration were identified. There are temporal differences in atmospheric particle pollution level, which is regarded as the boundary of the neighbourhood-scale pollution. Abundant research has reported that the local particle concentrations are related to the meteorological conditions, which directly influence their deposition processes (Jacob and Winner, 2009; Tai *et al.*, 2010; Tian and Chen, 2010). The urban form has an influence on the airflow (Z. Li *et al.*, 2019), which affects the dispersion of pollutants, and the vortex generated plays an important role in the retention of pollutants. There are also many sources of particle pollution in a city, such as traffic and construction sites. Transportation emits contaminants produced by the combustion of fossil fuels

181 (Fan *et al.*, 2018; Giovanis, 2018), whose contribution to total emissions into the air reaches 7.61%  
182 for PM<sub>10</sub> and 9.98% for PM<sub>2.5</sub> in Europe (European Environment Agency (EEA), 2018).  
183 Construction activities deteriorate air quality (Dong and Ng, 2015) in the process of land clearing,  
184 the operation of diesel diggers and generators, demolition, burning, mixing and so on (Zuo *et al.*,  
185 2017). These sources directly discharge pollutants to adjacent areas, resulting in an increased  
186 particle concentration with little timely diffusion. From the above analysis, four categories of data  
187 are required for modelling as predictors, which are described as follows:  
188

### 189 **(1) Background particle pollution level**

190 The local emission, dispersion and deposition status contributes to the overall air pollution  
191 level on a macro scale; in return, the local air pollution level can be considered using an overall air  
192 pollution level added to the features influencing the production and movement of pollutants. Hence,  
193 the particle pollution monitoring data from some official observation sites near the ground are used  
194 to represent the overall pollution level. This information is available on official measurement sites  
195 in the studied areas containing data from a number of scattered locations. It indicates the overall  
196 level of particle concentrations for the whole area at a particular time.  
197

### 198 **(2) Meteorological conditions**

199 Studies have shown that particle concentrations are related to meteorological variables. Tai *et al.*  
200 (2010) reported that the PM<sub>2.5</sub> concentration tends to be lower at high wind speeds, as wind force  
201 helps the dispersion of PM. Temperature is mostly found to be positively correlated with particle  
202 concentration (Tai *et al.*, 2010; Tian and Chen, 2010). Precipitation efficiently scavenges PM as  
203 with wet deposition, which makes it negatively related to particle concentration (Jacob and Winner,  
204 2009; Tai *et al.*, 2010). Therefore, the meteorological conditions around the target areas are essential  
205 parameters. The meteorological parameters including ground-level (2m height) air temperature,  
206 relative humidity, wind speed and precipitation are used as predictors in this research.  
207

### 208 **(3) Urban morphology**

The physical environment of cities as determined by dimensions, densities and infrastructure patterns, directly influences the configuration of the urban atmosphere and affects the urban microclimate and air contamination (Z. Li *et al.*, 2019). Urban morphology is an important consideration for urban planning, some categorized patterns are shown with neatly arranged urban structures (Ratti *et al.*, 2003). Given that the arrangement of buildings could be scattered and quite random, subject to the complicated topographical conditions, this research attempts to use some generalized indices to describe the building arrangement patterns. There are many factors used to describe urban morphology corresponding to different scales of interest. For the neighbourhood or block scale (0.1 ~ 10km) this research focuses on, the building coverage ratio (BCR), average building height (BH), building volume density (BVD) and the frontal area (FA) index are often used. There is evidence that the floor area ratio and building density are positively associated with particle concentrations in some cities (Shi *et al.*, 2019).

BCR is the percentage of the total area covered by buildings in a target area, indicating the horizontal compactness of the infrastructure, which is the most commonly used index for quantifying the building density at land lot scale (Yu *et al.*, 2010):

$$BCR = \frac{\sum_{i=1}^n A_i}{S}$$

(1)

where  $S$  is the total target land area;

$A_i$  is the coverage area of the building  $i$ ; and

$n$  is the total number of buildings in the target area.

BH here is coverage-area-weighted, i.e. the height of a building with a larger coverage area contributes more to the average building height of the target area:

$$BH = \frac{\sum_{i=1}^n (A_i \times h_i)}{\sum_{i=1}^n A_i}$$

(2)

where  $h_i$  is the height of the building  $i$ . This index shows the vertical extension of the land surface.

In this research, the BCR for different height levels (0m, 10m, 20m, 30m, 40m, 60m and 80m)

and the area-weighted average BH in a land lot of 500m\*500m are applied.

#### (4) Pollution sources

Industries, transportation and construction activities are recognised as the three main pollution sources in an urban area (Xu *et al.*, 2018). Assuming there is no polluting factory in the central urban area, the magnitudes of transportation and construction in each surveyed area are calculated using the metrics described below.

##### **Transportation:**

Roads are one of the pollution emission sources in an urban area (Health Effects Institute, 2010; Sun *et al.*, 2018). It is challenging to obtain real-time counts for the running flow of different types of vehicle. However, the statistics of transportation facilities and information from the real-time released platform of road condition can be used to represent the pollutant emission level of the locations.

Urban transportation infrastructure investment is related to air pollution (Sun *et al.*, 2018). The length of each road on a 500m\*500m buffer area centred on the sampling point can be measured, and the number of lanes for each road can be counted, hence the single-lane road length per unit area (SLRL) can be calculated using:

$$SLRL = \frac{\sum_{i=1}^m (L_i \times N_i)}{S}$$

(3)

where  $S$  is the total target land area,

$L_i$  is the length of the road  $i$ ;

$N_i$  is the number of lanes for the road  $i$ , and

$m$  is the total number of roads in the target area. The  $SLRL$  index shows the scale of road construction, reflecting the possible density of traffic pollution sources in the surrounding area.

For the direct influence of nearby pollution sources, the main road near the sampling point is selected, and its distance measured. The congestion status was accessed from the navigation software. The congestion status is categorized into four levels: i.e. green for ‘clear’, yellow for

‘slow-moving’, red for ‘congested’ and red-black for ‘heavily congested’, however, the specific vehicle velocities of each status depend on the road speed limits, which can also be obtained through field investigation. Finally, the distance to the nearest main road, with its speed limit, lane count and congestion status act as inputs into the model as the estimators for local traffic emissions.

#### **Construction activities:**

A large amount of dust generated from a construction site can spread over a wide area over a long period (Greater London Authority, 2014). The area of construction sites and the distance from the sampling point are input into the model as the estimators for construction emissions. If there is no construction site appearing in the surrounding area, the area of construction sites is set as 0m<sup>2</sup>, and the distance is set as 10km.

Table 1 lists all the predictors identified for the ANN model. The tick for ‘Temporal’ indicates the data varying with time, and the tick for ‘Spatial’ indicates the data varying with location. The day in a week (W = 1 for Monday, 2 for Tuesday... 7 for Sunday) and the hour in a day (hh = 0, 1, 2... 23) are also added into the predictors for capturing the law of periodic variations.

Table 1: The list of predictors used in the ANN model.

Categories	Predictors	Indices for input	Temporal	Spatial
Time periodicity	Week	$\sin(W/7*2\pi)$ and $\cos(W/7*2\pi)$	√	
	Hour	$\sin(hh/24*2\pi)$ and $\cos(hh/24*2\pi)$	√	
Background level	Monitoring from regulatory sites (μm.m <sup>-3</sup> )	Average $PM_{10}$ or $PM_{2.5}$ concentrations	√	
Meteorological conditions	Temperature (°C)	$Temp$	√	√
	Relative humidity (%)	$RH$	√	√
	Wind speed (m s <sup>-1</sup> )	$WS$	√	
	Precipitation (mm)	$RF$	√	
Urban morphology	BCR for different height levels in a land lot of 500m * 500m	$BCR_0, BCR_{10}, BCR_{20}, BCR_{30}, BCR_{40}, BCR_{60}$ and $BCR_{80}$		√
	BH in a land lot of 500m*500m (m)	$BH$		√

Categories		Predictors	Indices for input	Temporal	Spatial
Pollution sources	Emissions from traffic in the local area.	Distance to the nearest main road (m)	$D_t$		√
		Speed limit of the nearest main road (km.h <sup>-1</sup> )	$SL_t$		√
		Lanes count of the nearest main road	$LC_t$		√
		Average congestion status in a land lot of 500m*500m (0, 1.0~4.0)	$CS_t$	√	√
	Emissions from traffic in the surrounding area.	Single-lane road length per unit area in a land lot of 500m*500m (km.km <sup>-2</sup> ).	$SLRL$		√
	Emissions from construction activities.	Area of construction site within 500m (m <sup>2</sup> ).	$A_{cs}$		√
		Distance of nearest construction site (m).	$D_{cs}$		√

279

## 280 2.2 Field measurements for particles (*Step 2*)

281 In this step, locations are to be selected for the measurements of particle concentrations, and  
 282 the real-time measured value at the specific location represents the predicted variable. One of the  
 283 feasible field measurement procedures is depicted in the case study example (Section 3.1).

284

## 285 2.3 Data-driven modelling and verification (*Step 3*)

286 ANN-based, data-driven modelling is an entirely different approach to conventional  
 287 algorithms. It is normally a computing system vaguely inspired by the biological neural networks  
 288 that constitute human brains (Haykin, 2009). The structure of a fully connected feed-forward ANN  
 289 consists of the input layer, the hidden layers and the output layer (Figure 2-a). The activation of  $a_j^l$   
 290 (the  $j^{\text{th}}$  neuron in the  $l^{\text{th}}$  layer) is related to the neurons in the  $(l-1)^{\text{th}}$  layer (Figure 2-b) by the  
 291 equation:

$$292 \quad a_j^l = f\left(\sum_{k=1}^{n_{l-1}} w_{jk}^l a_k^{l-1} + b_j^l\right) \quad (4)$$

293

294 where  $a_k^{l-1}$  is the  $k^{\text{th}}$  neuron in the  $(l-1)^{\text{th}}$  layer;



295  $n_{l-1}$  is the total number of neurons in the  $(l-1)^{\text{th}}$  layer;

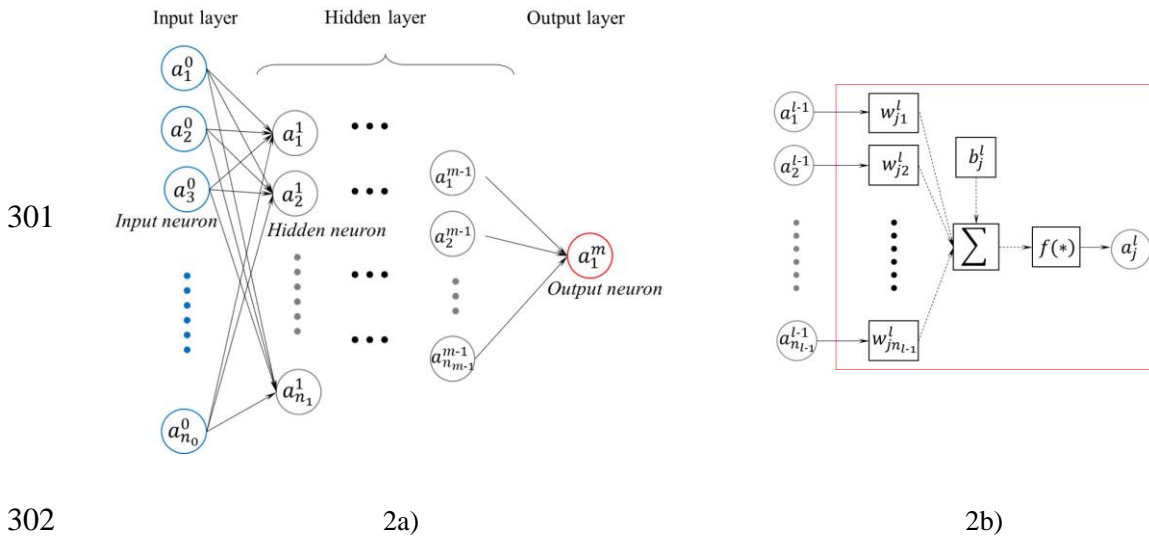
296  $w_{jk}^l$  is the weight for the connection from the  $k^{\text{th}}$  neuron in the  $(l-1)^{\text{th}}$  layer to the  $j^{\text{th}}$  neuron in the  $l^{\text{th}}$

297 layer;

298  $b_j^l$  is the bias of the  $j^{\text{th}}$  neuron in the  $l^{\text{th}}$  layer; and

299  $f(*)$  is the activation function, which determines its nonlinear properties.

300



303 Figure 2: The structure of fully-connected feed-forward ANN. a) The whole network structure; b) The internal  
 304 operations of a neuron.

305

306 The package “caret” (Kuhn., 2018) in the software R (v 3.5.1) (R Core Team, 2018) is used to

307 train the ANN model. All the data for the predictors are fed into the input neurons and the

308 measurement data are fed into the output neuron. The whole dataset is randomly divided into two

309 subsets, one for model training and the other for testing using the ratio of 3:1. The cross-validation

310 is used in the training process using the training dataset. The testing dataset is individually used for

311 the verification of the ANN model.

312 The effectiveness of the prediction can be evaluated by statistics measuring how well the

313 observed outcomes are replicated by the model. The *root mean square error (RMSE)* and the *mean*

314 *absolute error (MAE)* are the most common indicators used with prediction models. *RMSE* uses the

square root of the second sample moment of the differences between predicted values and measured values to represent the overall accuracy, i.e.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (P_i - M_i)^2}{n}}$$

(5)

where  $P_i$  is the  $i^{\text{th}}$  predicted value,  $M_i$  is the  $i^{\text{th}}$  measured value, and  $n$  is the volume of the datasets to compare.

The Pearson correlation coefficient ( $r$ ), a value between -1 and +1, is a measure of the linear correlation between predicted values and measured values, i.e.

$$r = \frac{\sum_{i=1}^n (P_i - \bar{P})(M_i - \bar{M})}{\sqrt{\sum_{i=1}^n (P_i - \bar{P})^2} \sqrt{\sum_{i=1}^n (M_i - \bar{M})^2}}$$

(6)

where  $\bar{M}$  is the average of the measured values, and  $\bar{P}$  is the average of the predicted values.

The average bias (*Bias*), or say the average of the predicting errors, is calculated to describe how much the model underestimates or overestimates the situation, thus:

$$Bias = \frac{\sum_{i=1}^n (P_i - M_i)}{n}$$

(7)

Relative error histograms are plotted to show the frequency of the appearance of errors at a different scale, which tells what percentage of the data lies within the acceptable tolerance.

## 2.4 Application for estimation – Spatial interpolation (Step 4)

After training and verification of the model, it would be theoretically possible for the estimation of particle concentrations at any location and time, as long as all the information for the prediction variables is provided. Thus, one of its applications could be a spatial interpolation.

An area of interest can be divided into a 500m\*500m grid. All the data for the predictors with

spatial variations ( $BCR$ ,  $BH$ ,  $SLRL$ ,  $CS_t$ ,  $D_t$ ,  $A_{cs}$  and  $D_{cs}$ ) are calculated with the aid of GIS and self-developed Python scripts.

In general, spatial-variable factors could be divided into two types: building information and road information. The building information, as vector data, could be used for spatial analysis. However, road information is in the form of raster data (like an image) which should be converted into vector data. In order to extract useful information from road information data and convert it to the vector data format, the ModelBuilder, which could be thought as a visual programming language application in ArcMap (a GIS program), is applied to process the data. Figure 3 is the work chart for extracting road data in the ModelBuilder. In addition, the extracted road information could be converted into vector data for spatial analysis. After obtaining construction and road spatial data in vector format, a fishnet, namely dividing an area into finite small squares, is used to count spatial features at different locations.

350

351

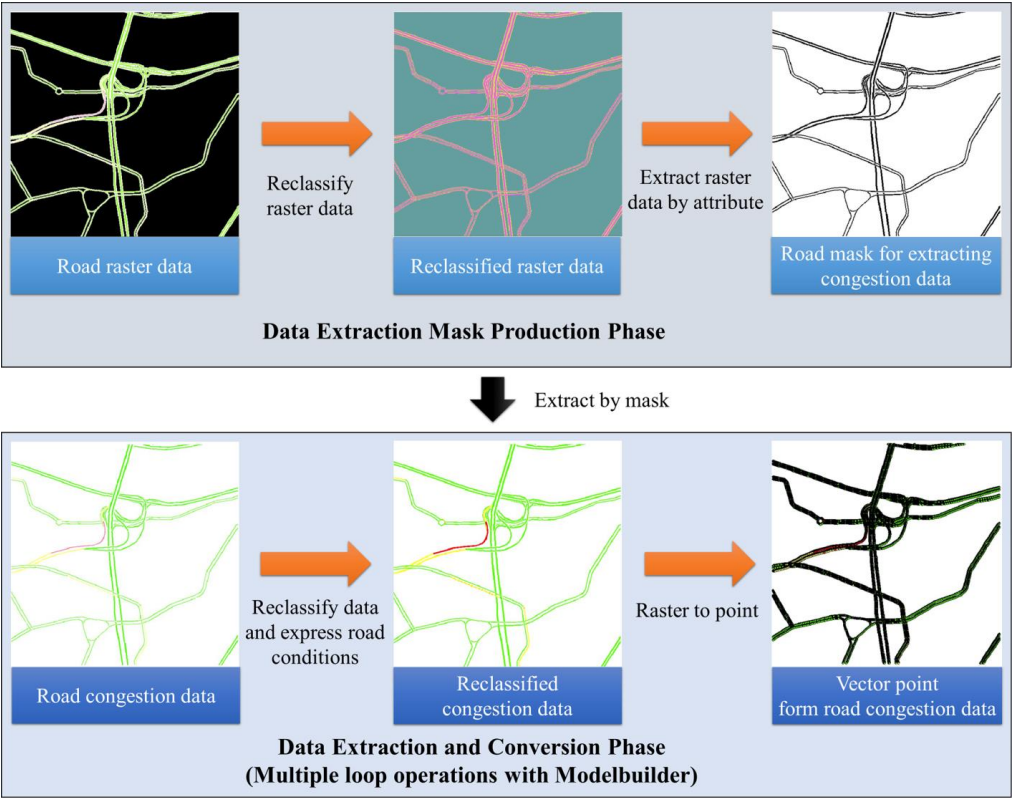


Figure 3: Flow chart for extracting road information.

353

354

355

356

357

358

359

360

361

362

363

364

365

366

367

368

369

370

371

372

373

374

375

376

377

378

In order to calculate these spatial variables, the ‘Spatial Join’ (Esri., 2019a) and ‘Near’ (Esri., 2019b) in the Analysis tools of the ArcMap are mainly used. The Spatial Join is the tool used to connect the properties of one feature class to the properties of another feature class, based on spatial relationships. To be more specific, this tool could be used to calculate the length of the road, the total area and the number of buildings in a region. Hence, spatial variables of  $BCR$  at different heights,  $BH$ ,  $SLRL$ ,  $CS_t$  and  $A_{cs}$  are calculated through the Spatial Join tool in the ArcMap. Additionally, the Near tool is used to calculate the distance and other proximity information between the input features and the closest features in other layers or feature classes. Therefore, the spatial information for  $D_t$  and  $D_{cs}$  is analysed by the Near tool in the GIS software.

The corresponding data for each predicting variable for every grid forms a dataset, which is input into the trained model, and the output is the corresponding particle concentrations of each grid location.

### 3 Verification of the method using a case study

Chongqing has become one of the fastest developing cities in China, accompanied by rapid urbanisation and infrastructure construction on a grand scale. Consequently, its ambient air quality has been gradually deteriorating over the last few years. Chongqing was selected as the case-study city in this research to verify and demonstrate the process for developing this research method and its application.

#### 3.1 Measurement of real-time particle concentrations

The data used for this study was from field measurements carried out in the dense central urban area of Chongqing between July 2015 and January 2016 covering summer, autumn and winter seasons. For security reasons, monitoring devices were located in some residences, and the sampling tube was extended out of the window with a pole. A total of 20 dwellings was selected in

central districts (Figure 4). Continuous 4~5 days monitoring data were collected for each location successively (totally 84 days). The field measurement period for each location is indicated in the Supplementary Material 1.

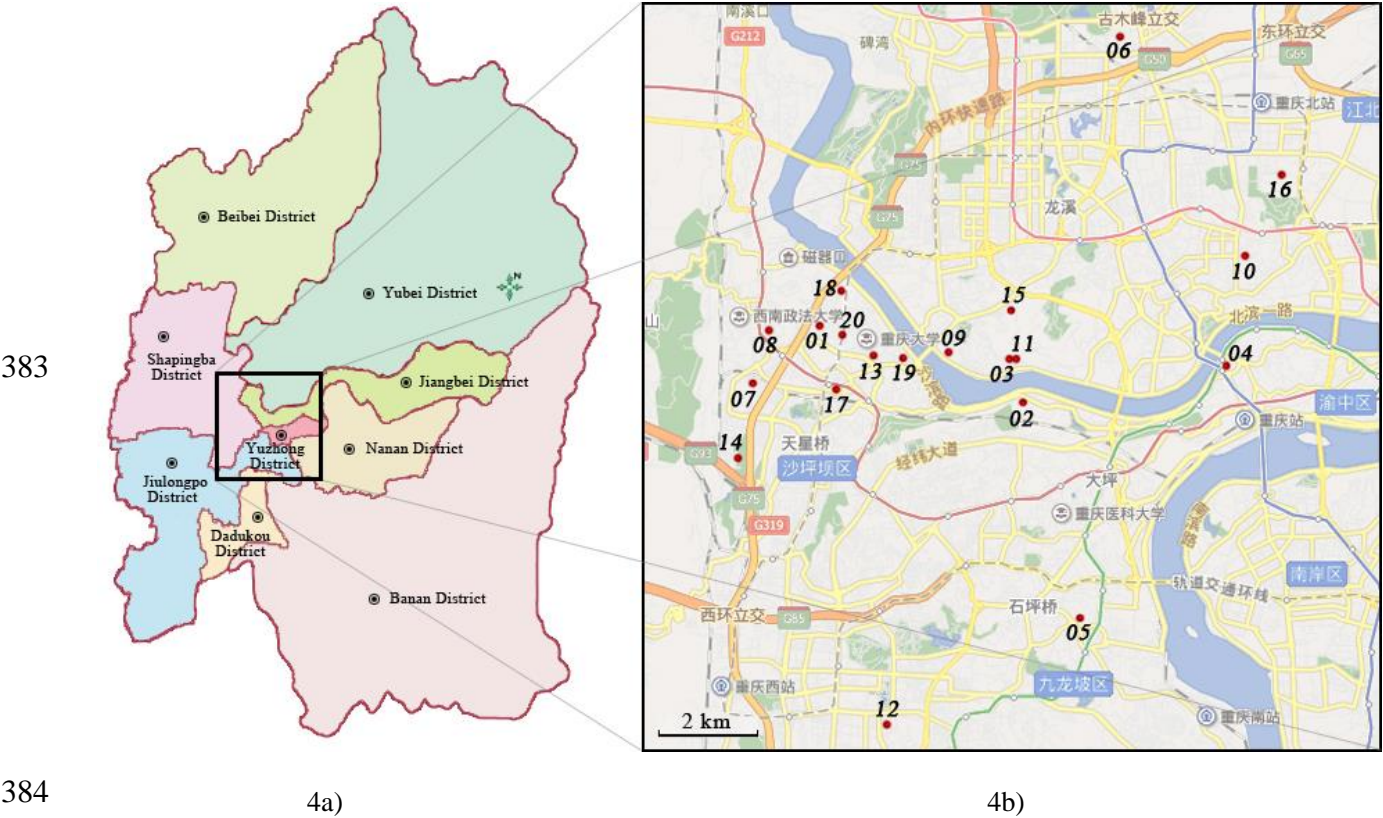


Figure 4: The location of the field measurement campaign. a) The central urban area of Chongqing (the black square frame); b) The distribution of sampling sites (red dots).

The measured parameters include temperature, relative humidity and PM concentration (Table 2). In order to measure these parameters accurately, avoiding the influence of indoor disturbances, the sampling point was located 2 metres outside the window or balcony, and a supporting rod was specially laid for this purpose (Figure 5). Onset HOBO UX100-011 is an automatic logger comprising a temperature sensor, an RH sensor and memory to record the data. It was directly hung on the end of the rod due to its small size. TSI DustTrak 8534 is a light-scattering laser photometer that gives real-time aerosol mass readings, which can simultaneously measure size-segregated mass

fraction concentrations corresponding to  $PM_{2.5}$  and  $PM_{10}$ . This device uses a sheath air system that isolates the aerosol in the optics chamber to keep the optics clean for improved reliability and low maintenance. Jiang (Jiang, 2013) has conducted a series of experiments to verify that the results from the aerosol monitoring method using DustTrak DRX have strong consistency with the results from a tapered element oscillating microbalance. It was calibrated with the zero filters every day before the sampling started. All the monitoring equipment was set-up to log data at 1-min intervals, and the collected data could be readily processed for specific purposes.

Table 2: Real-time measuring equipment for temperature, relative humidity, PM concentration ( $PM_{10}$  and  $PM_{2.5}$ ), and their technical specifications.

Model	Manufacturer	Variables	Range	Accuracy	Resolution
HOBO UX100-011	Onset	Temperature	-20 ~ 70 °C	± 0.21 °C (0 ~ 50 °C)	0.024 °C
		Relative humidity	1% ~ 95%	± 2.5% (10% ~ 90%) ~ ± 3.5% (0% and 100%)	0.05% (25 °C)
DustTrak 8534	TSI	PM concentration	0.001 ~ 150 mg m <sup>-3</sup>	± 0.1% of reading	0.001 mg m <sup>-3</sup>



Figure 5: Measurement devices for outdoor thermal conditions and particle pollution levels. a) The location of outdoor sampling point; b) The measuring equipment.

410

## 411 3.2 The dataset for the predictors

### 412 3.2.1 Background pollution level

413 The hourly PM<sub>10</sub> and PM<sub>2.5</sub> data are obtained from the National Air Quality Real-time Release  
414 Platform (<http://106.37.208.233:20035/>) (China National Environmental Monitoring Centre) by the  
415 China National Environmental Monitoring Centre. There are 6 official observation sites (with  
416 reference numbers ‘1414A’, ‘1417A’, ‘1419A’, ‘1423A’, ‘1424A’, and ‘1425A’) in the central  
417 Chongqing area selected for the case study, and an average of 6 sites made up the predicting dataset.

418 From the particle monitoring data in the official observation sites (Figure 6), we can see that  
419 the most severely polluted days are aggregated in winter, but there is still a lot of time in other  
420 seasons that have reached the limit. However, the limit set by the Chinese government(General  
421 Administration of Quality Supervision, Inspection and Quarantine and China, 2012), which is  
422 150 $\mu\text{g.m}^{-3}$  for PM<sub>10</sub> and 75  $\mu\text{g.m}^{-3}$  for PM<sub>2.5</sub> (red solid threshold line in Figure 6), is more relaxed  
423 than that of the World Health Organization (WHO) (2006) values, which is 50  $\mu\text{g.m}^{-3}$  for PM<sub>10</sub> and  
424 25  $\mu\text{g.m}^{-3}$  for PM<sub>2.5</sub> (red dotted threshold line in Figure 6); consequently, most of the days cannot be  
425 regarded as a “safe day” when compared to the WHO standard values.

426



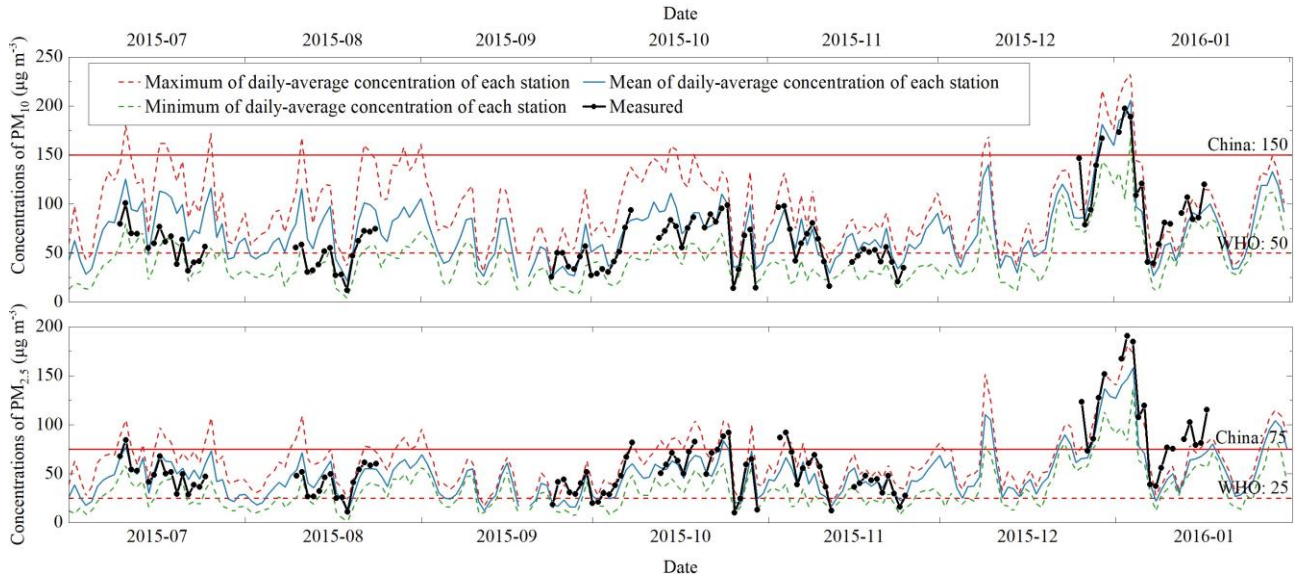


Figure 6: The 24-h average particle concentrations: a)  $PM_{10}$ ; b)  $PM_{2.5}$ . The data from surrounding air quality monitoring sites: blue for the average, red for the maximum, and green for the minimum, and the data from the field measurement: the black dotted line.

The on-site measurements of  $PM_{10}$  and  $PM_{2.5}$  are compared with the officially released data (Figure 6). A similar trend is observed for the PM concentration throughout the urban area of Chongqing. However, the pollution level varies for different regions within urban areas, which indicates the importance of the spatial interpolation of pollution levels in obtaining local pollution status.

### 3.2.2 Meteorological conditions

Daily and hourly weather observations are obtained from the China Meteorological Administration (<http://data.cma.cn/>) (China Meteorological Administration). The observation site chosen is called Shapingba (57516), which is located in the urban area of Chongqing, and it is the closest to all the on-site measuring points.

The entire measurement period spanning from summer through autumn to winter, experiences all kinds of typical climate conditions for Chongqing (Figure 7). This city suffers a continuous



heatwave from the beginning of July to the beginning of September with an average temperature of 28.4°C, and there were totally 21 days when the highest temperature reaches 35°C from 7<sup>th</sup> Jul. to 10<sup>th</sup> Sep. 2015, with the daily lowest temperature peaking at 29.3°C on 2<sup>nd</sup> and 3<sup>rd</sup> Aug. 2015. Thereafter, a warm-season lasted for 1.5 months from 11<sup>th</sup> Sep. to 25<sup>th</sup> Oct. 2015 with an average temperature of 21.8°C. The autumn in Chongqing is very short from the end of October for one month, declining sharply towards early winter with the air temperature averaging 9.2°C from 13<sup>th</sup> Dec. 2015 to 20<sup>th</sup> Jan. 2016. The humidity is high throughout the year, with an average relative humidity of 77.7%, and there are 89 days when the average relative humidity is above 80% (1<sup>st</sup> Jul. 2015 – 31<sup>st</sup> Jan. 2016). Chongqing is categorised in the calm wind zone with an average wind speed of around 1m.s<sup>-1</sup>. In that summer, most of the days were exposed to sunlight, except several days (15<sup>th</sup> and 22<sup>nd</sup> Jul., 17<sup>th</sup> Aug., 5<sup>th</sup> and 12<sup>th</sup> Sep. 2015) with rainstorms (>50mm in 24 hours). However, the sunlight is very rare for this region in winter when most days are very humid with drizzle.



Figure 7: The weather conditions during the measurement period. a) Temperature, including daily mean, maximum and minimum value from weather station (line chart), and statistics from the field measurement (boxplot); b) Relative humidity, including daily mean and minimum values from weather stations (lines chart), and statistics from the field measurement (boxplot); c) Wind speed, including daily maximum and mean values; d) Sunshine hours, total hours of sunny time in a full day; e) Precipitation, total rainfall in 24 hours (last night 20:00 to 20:00).

The black dots with IQR bar (Figure 7) show the measurement of temperature and relative humidity from the field tests. It follows the trend captured by the weather stations. For the context of the urban environment, the urban heat island effect makes the positive bias (+ 0.98 °C) for almost all the temperature measurements. The highest local temperature during the period of the field

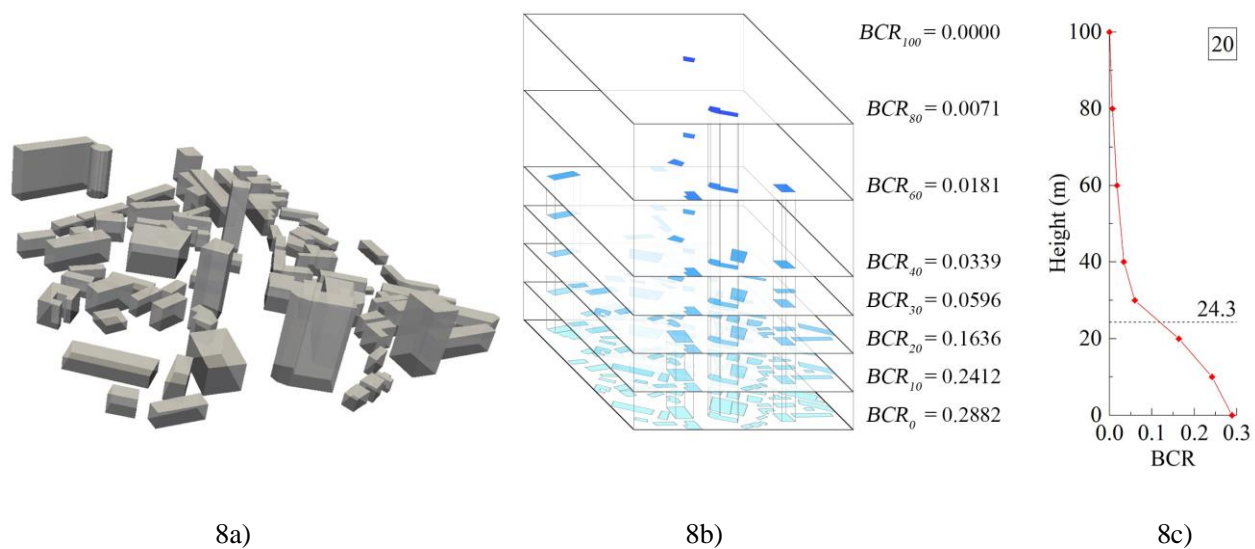
471 measurement reached 42.6°C (15:00 12<sup>th</sup> Aug. 2015).

472

473 **3.2.3 Urban morphology**

474 The BCR at different heights (Figure 8-b) is calculated to express the urban form for the density  
475 of the buildings (which reflect the changes in the vertical direction), using a set of values to depict  
476 more details of the three-dimensional morphological characteristics of the urban area. Furthermore,  
477 the building volume per unit land area could be estimated by the area enclosed by the polyline and  
478 the coordinate axis (Figure 8-c). In general, the BCR at different heights and BH are not exhaustive  
479 but sufficient enough to reflect the impact of urban morphology on the dispersion of air pollutants  
480 in this research.

481



482

483

484 Figure 8: Numerical transformation of urban morphology. a) The actual building model of the location ‘20’; b)  
485 The BCR for different height levels on this land plot; c) The dotted line diagram of the relationship between BCR  
486 and height level, and the dashed line indicates the BH of this land plot.

487

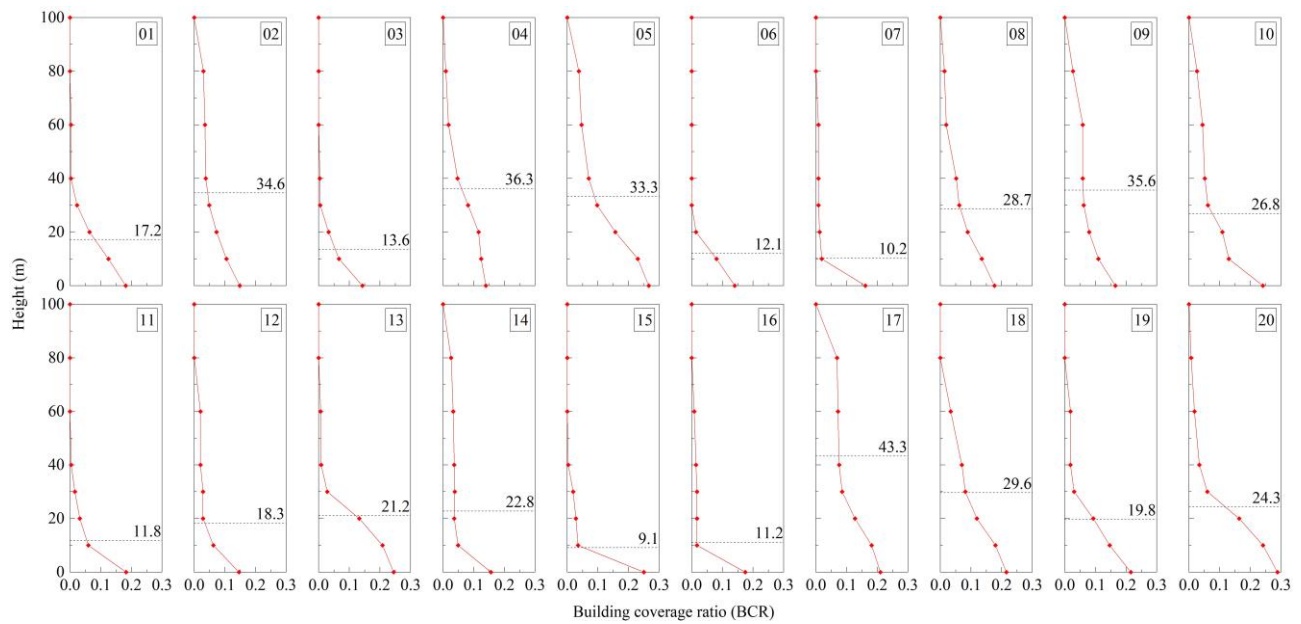
488 The BCR at different height levels and the BH are calculated as the urban morphology  
489 characteristics of input variables (Figure 9). Given that government regulations impose no

490 restrictions on building height in Chongqing, both high and low buildings are found together in the  
 491 central urban area. The highest building in these surveyed areas is lower than 100 metres. Buildings  
 492 in the non-commercial area generally meet this rule because the super-high-rise buildings (greater  
 493 than 100 meters in height) need to follow a much stricter design and construction code. (The  
 494 Ministry of Housing and Urban-Rural Development of the People's Republic of China, 2005).

495 The selected areas in this study have different morphological characteristics. For example, the  
 496 lowest ground-level density is 0.1393 at location '06', and the highest is 0.2882 at location '20'.  
 497 Almost no high-rise buildings are shown in locations '01', '03', '06', '11', '13' and '15'; high-rise  
 498 buildings are very sparsely present in locations '04', '07', '08', '12', '16', '19' and '20' but appear  
 499 more frequently in locations '02', '05', '09', '10', '14', '17' and '18'.

500

501



502 Figure 9: The BCR on each measurement point for different height levels (the dashed line indicates the BH in that  
 503 area).

504

### 505 3.2.4 Local pollution sources

#### 506 1) Transportation

The transportation facilities were identified using the satellite image provided by the software Google Earth Pro (version 7.3.2) on 21<sup>st</sup> Oct. 2015, which was during the field measurement period (Figure 10-a). The congestion status was accessed from the navigation software Baidu Map (<https://map.baidu.com/>) at around half-hourly intervals (Figure 10-b).

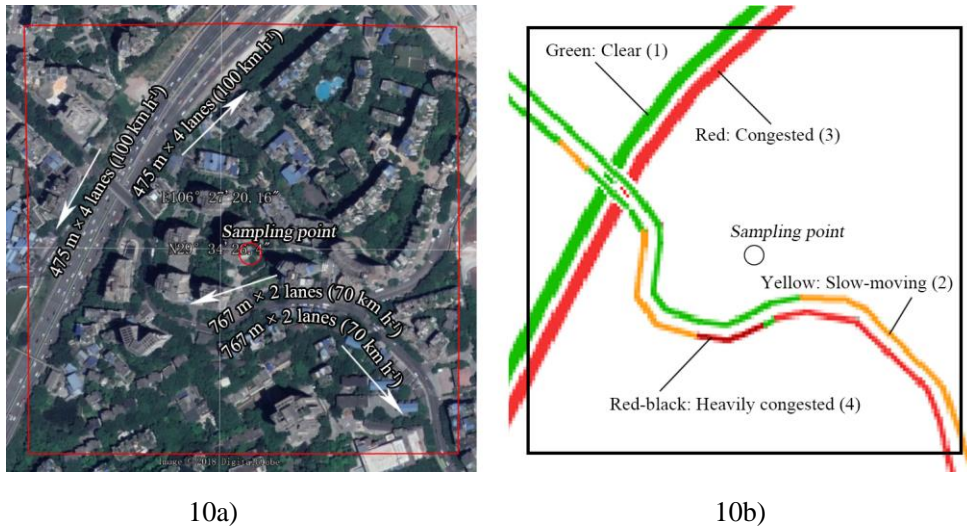


Figure 10: Traffic information around the sampling point. a) The satellite image of location '18'; b) The congestion status.

All the variables providing information on emissions are dependent on the locations (see Supplementary Material 1). The single-lane road length per unit area (SLRL), indicating the density of road facilities, varies from  $7.9 \text{ km.km}^{-2}$  (a relatively isolated residential community) to  $37.3 \text{ km.km}^{-2}$  (entrance of an inner-ring highway) with an average of  $22.11 \text{ km.km}^{-2}$  (standard deviation:  $7.96 \text{ km.km}^{-2}$ ).

The temporal variations of traffic emissions are characterised by the time periodicity and the congestion status (Figure 11). For weekdays, the roads used for work commuting generally have two distinct peaks, which appear in the residential, the commercial for offices and schools, and the inner-ring highway areas. However, around the commercial areas for entertaining and shopping, the traffic conditions are not smooth for the whole day. For the weekend, the urban traffic congestion profile is more diverse. It was smooth for the whole day in the residential areas and the commercial

areas for offices and schools. A peak shows in the afternoon due to a sudden intense utilization of the highway. The road around the commercial areas for entertaining is congested almost the whole day, even worse than that during a weekday, and a peak appears at night towards the end of non-home-based activities. This information reflects the road usage at different times and indirectly supports the estimation of traffic pollutant emissions.

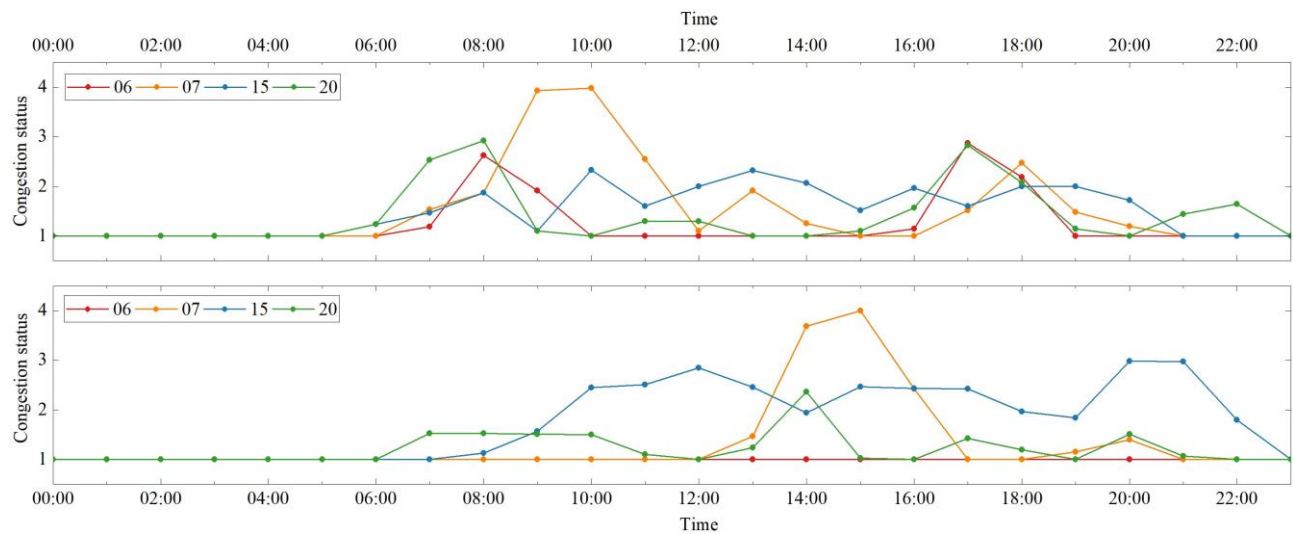


Figure 11: Real-time congestion status for a) a weekday and b) a weekend on four typical locations: '06' (Road in the residential area), '07' (Inner-ring highway), '15' (Road in the commercial area for entertaining) and '20' (Road in the commercial area for offices and schools). Congestion status was interpolated from four-level road conditions as shown in Figure 10.

## 2) Construction activities

A construction site was identified within 500m of the sampling point using the satellite image provided by Google Earth Pro software on 21<sup>st</sup> Oct. 2015 and its area and distance from the measurement point obtained. Assuming the construction sites have not changed during the period of field measurements, the data for these two variables are constant and calculated as shown in Supplementary Material 1. There was a lot of construction work during that time due to intense development. 16 locations (out of 20) appeared the construction sites, the largest of which was 190

metres away from the test point with an area of around 127,437m<sup>2</sup>.

### 3.3 Predicted results and verification

The whole dataset is prepared following the above instructions and provided in Supplementary Material 2. The ANN scripts are provided in Supplementary Material 3. Based on the comparison with the testing dataset, the predicted results from the ANN model with background pollution level, weather conditions, urban morphology and local pollution sources are in good agreement with the measured data (Figure 12). A linear relationship between predicted values and measured values is found with a Pearson coefficient of 0.954 for PM<sub>10</sub> (sig. <0.001), and 0.968 for PM<sub>2.5</sub> (sig. <0.001). The mean square error for PM<sub>10</sub> is 11.20µg.m<sup>-3</sup>, and 9.04µg.m<sup>-3</sup> for PM<sub>2.5</sub>. The bias is +1.07µg.m<sup>-3</sup> for PM<sub>10</sub>, and +0.98µg.m<sup>-3</sup> for PM<sub>2.5</sub>. However, when observing the data in Figure 12, the positive errors appear for the higher concentrations with the negative bias mainly being seen for lower concentrations.

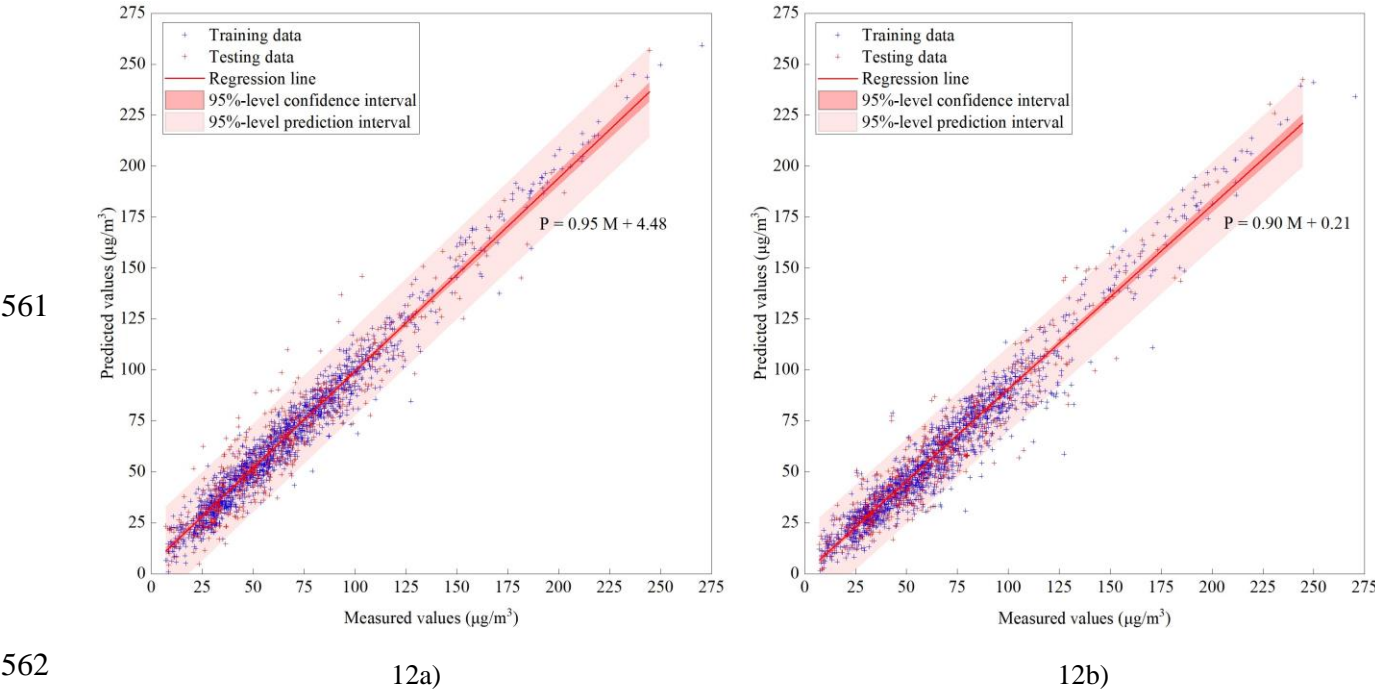


Figure 12: The comparison between predicted values and measured values of a) PM<sub>10</sub> and b) PM<sub>2.5</sub>.



564

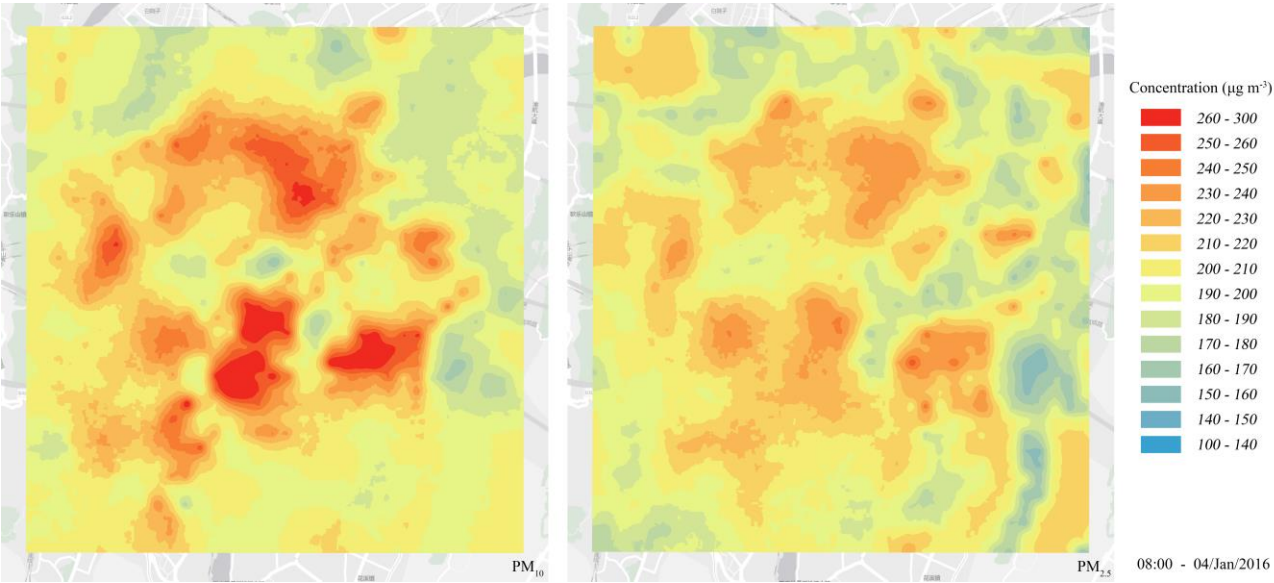
565 **3.4 Application for spatial interpolation**

566 As the data-driven prediction model has been developed for the studied area, the particle  
567 concentrations can be estimated at a specified time and place for this area. To obtain the average  
568 particle concentrations for 6 official sites, the meteorological parameters can be accessed from the  
569 officially released platform at the given time. Information on the urban morphology and local  
570 sources can be processed using satellite images and the GIS system. Then all the data for the  
571 predictors are required to be fed into the model which then outputs the predicted concentration  
572 values.

573 Following the instructions in Section 2.4, the concentrations of  $PM_{10}$  and  $PM_{2.5}$  in each  
574 500m\*500m grid at 08:00 on 04 Jan 2016 are estimated. The mapping of the concentration  
575 distribution (Figure 13) is smoothed out by the Empirical Bayesian Kriging method (Esri, 2018).  
576 The centre is a more densely built area with a greater population than its surroundings, and the  
577 traffic flow is also high, hence it is not surprising to find that the PM concentrations are higher at  
578 the centre of this image.

579

580



581

13a)

13b)



582 Figure 13: The prediction of a)  $PM_{10}$  and b)  $PM_{2.5}$  concentrations of the whole case study area at 08:00 on 04 Jan  
583 2016.

584

585 This model can also be used in any other urban area. Some typical sites need to be selected to  
586 conduct real-time particle monitoring. The particle pollution monitoring data from official  
587 observation sites can be accessed from the local authorised sources. The meteorological conditions  
588 can be accessed from the local meteorology department. For the urban form, the urban planning  
589 department may have such information, however, it also can be obtained by satellite images, and the  
590 related indices can be processed according to the method described above. The road transportation  
591 infrastructure and construction sites can be read from satellite images, and the traffic conditions can  
592 be accessed from the contemporary navigation system. With all the information obtained, the  
593 predicted values can be calculated using the trained and validated ANN model.

594

## 595 **4 Discussion**

### 596 **4.1 Sensitivity analyses**

597 The prediction accuracy of the trained model is largely influenced by the dataset for training  
598 and testing. Factors of influence include, but are not limited to, the selection of the predictors, the  
599 volume of the data set and whether the training data cover the possible span of the predictors. The  
600 following sections discuss two of the issues that affect the accuracy of the model.

601

#### 602 **1) The influence of different combinations of predictors**

603 In this study, as is discussed in Section 1.1, five elements are considered as predictors in the  
604 model: time periodicity, background pollution level, weather conditions, urban morphology and  
605 local pollution sources, see Table 3. The trained ANN model is a spatial interpolation model  
606 considering the local divergence denoted as SC0. In order to test the impact of the number of  
607 predictors on the modelling accuracy, we tested another two cases. SC1 is the case which omits the

predictor for the background pollution level, and SC2 is the case which omits the predictor for ‘urban morphology’.

Table 3: Three input variable schemes are considered from the literature review for comparison.

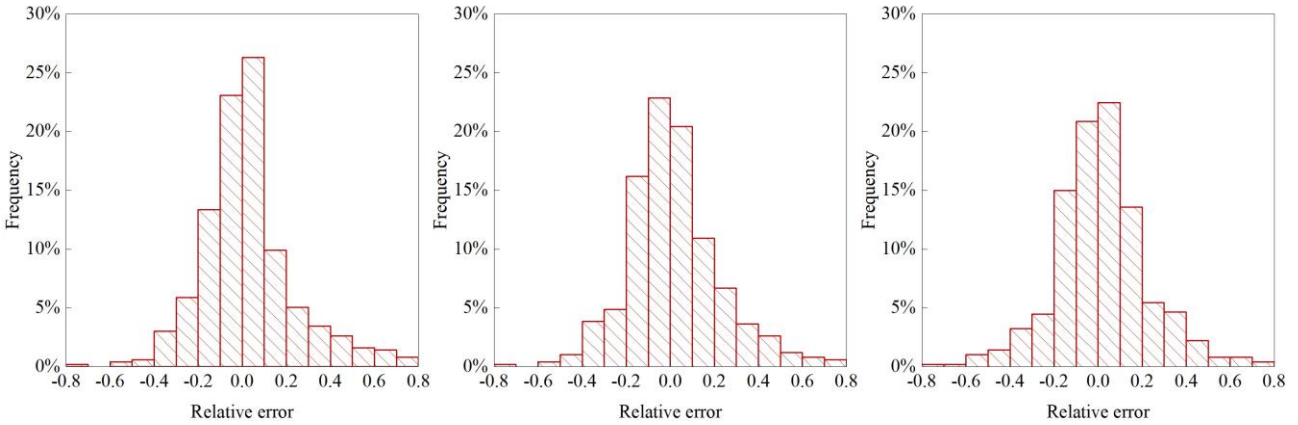
Input variable scheme	Categories				
	Time periodicity	Background particle pollution level	Meteorological conditions	Urban morphology	Pollution sources
SC0	√	√	√	√	√
SC1	√		√	√	√
SC2	√	√	√		√

The predicted performances are presented in Table 4 and Figure 14. From the figure, we can see that the most accurate model is the one considering five predictors (SC0), which is discussed in the above-mentioned section. The other two cases also demonstrated a very good performance in prediction. The SC1 scheme has a Pearson coefficient of 0.938 for  $PM_{10}$  and 0.925 for  $PM_{2.5}$ . This input scheme can be used to predict the pollution level when there is no available information on real-time pollutant concentration in certain surrounding locations. The SC2 scheme has the worst performance in terms of presentation accuracy as it ignores the urban morphological information, unlike the other two schemes. Figure 14 shows the distribution of relative error of  $PM_{10}$  and  $PM_{2.5}$  respectively using the predicted value compared with the measured value. The relative error is most concentrated around 0 for SC0 but widely scattered for SC2.

Table 4: The statistics for the prediction performance of models with different predicting variable schemes (Table 3) compared with field measurements ( $n_{\text{test}} = 494$ ).

Predicting variable scheme	Prediction for $PM_{10}$				Prediction for $PM_{2.5}$			
	<i>RMSE</i> ( $\mu\text{g.m}^{-3}$ )	<i>r</i>	<i>Bias</i> ( $\mu\text{g.m}^{-3}$ )	<i>Average relative error</i>	<i>RMSE</i> ( $\mu\text{g.m}^{-3}$ )	<i>r</i>	<i>Bias</i> ( $\mu\text{g.m}^{-3}$ )	<i>Average relative error</i>
SC0	11.20	0.954	1.07	17.56%	9.04	0.968	0.98	16.04%
SC1	13.89	0.938	1.10	20.59%	13.67	0.925	1.30	21.13%

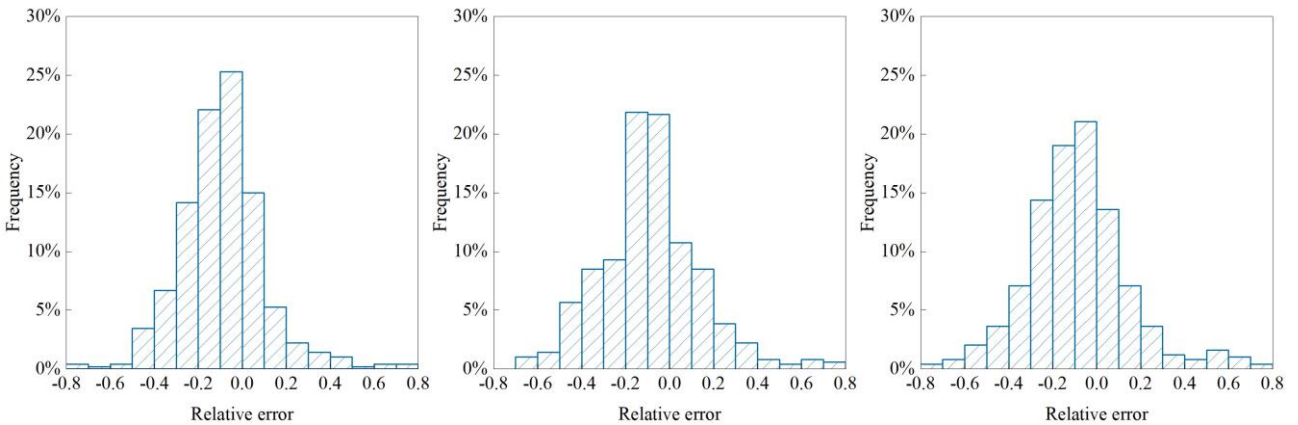
Predicting variable scheme	Prediction for PM <sub>10</sub>				Prediction for PM <sub>2.5</sub>			
	<i>RMSE</i> ( $\mu\text{g.m}^{-3}$ )	<i>r</i>	<i>Bias</i> ( $\mu\text{g.m}^{-3}$ )	<i>Average</i> <i>relative</i> <i>error</i>	<i>RMSE</i> ( $\mu\text{g.m}^{-3}$ )	<i>r</i>	<i>Bias</i> ( $\mu\text{g.m}^{-3}$ )	<i>Average</i> <i>relative</i> <i>error</i>
SC2	16.47	0.901	1.28	24.49%	15.50	0.896	1.41	24.06%



14a-1)

14a-2)

14a-3)



14b-1)

14b-2)

14b-3)

Figure 14: The histogram of the relative errors of a) PM<sub>10</sub> and b) PM<sub>2.5</sub> from models with different input variable schemes: a-1)/b-1) SC0; a-2)/b-2) SC1 and a-3)/b-3) SC2 (Table 3).

## 2) The influence of location selection

The locations used to train the prediction model will affect its accuracy. The ANN model was trained with data from 5, 10 and 15 locations respectively, and the accuracy of the model prediction

is given in Table 5. The results with 20 locations show a clear predictive power for the model, even though 20 locations may not be ideal, it is acceptable. The more locations are chosen, the more information about the urban morphology the model can learn, and the better its ability to predict other locations. Generally, the selection of locations should ensure the diversity of spatial morphologies in different locations.

Table 5: The effect of the number of selected locations on the accuracy of the model prediction.

Number of locations	$n_{\text{test}}$	Average relative error	
		PM <sub>10</sub>	PM <sub>2.5</sub>
20 locations (No. 01 - 20)	494	17.56%	16.04%
15 locations (No. 06 - 20)	379	19.50%	18.65%
10 locations (No. 11 - 20)	244	19.88%	19.48%
5 locations (No. 16 - 20)	119	23.28%	22.65%

## 4.2 Limitations and prospects

The application of the model is based upon the availability of predicting variables. Nowadays, these data are usually available in major cities worldwide provided by the local meteorological and air pollution observation stations. However, the application of the model is limited in regions that lack observation stations. Difficulties often arise in the acquisition of geographic information such as urban morphology and transportation networks, and their presentation forms vary from place to place, leading to the need to establish different data pre-processing schemes, as described in Step 1.

Subsequent studies will focus on the application of the model in other cities to demonstrate the applicability of the model worldwide.

## 5 Conclusions

This paper presents a newly developed holistic approach to predicting real-time urban particle concentrations in conjunction with spatial and traffic information datasets. Four variables are identified by considering the process of particle dispersion in the urban canopy layer: background particle concentrations, meteorological conditions, urban morphology and urban pollution sources.

659 The method of acquiring building and road traffic information has been developed by using GIS  
660 data, obtained from the urban planning information and satellite images, and self-developed Python  
661 scripts. The prediction model has been verified by a case study of Chongqing city. Continuous four-  
662 day measurements of  $PM_{10}$  and  $PM_{2.5}$  were conducted in 20 locations within the city centre area of  
663 Chongqing. The trained model has been verified with the results so that the average relative error of  
664 estimation compared with measurement was 17.56% for  $PM_{10}$  and 16.04% for  $PM_{2.5}$  showing the  
665 modelling to have a good degree of accuracy.

666 Sensitivity analysis has been conducted in order to test the accuracy level in the absence of the  
667 *background particle pollution level* or *urban morphology information*. The results show that the  
668 accuracy levels drop in both cases. For the former case, the relative errors dropped to 20.59% for  
669  $PM_{10}$  and 21.13 for  $PM_{2.5}$ . For the latter case, the relative errors dropped to 24.49% for  $PM_{10}$  and  
670 24.06% for  $PM_{2.5}$ . Sensitivity tests have also been done to examine the impact of the number of  
671 locations selected. It is obvious that the greater the number of locations selected, the more accurate  
672 the predicted pollution level is. The worse scenario of 5 locations will reach a relative error of  
673 22.65%.

674 The model is robust which suggests that it can be used in other cities with the required input  
675 parameters from local sources. It can serve as a tool for a fast estimation of particle concentration in  
676 an urban environment after the input of real-time information including particle concentration  
677 monitoring and meteorological observations from an official site, urban satellite images and traffic  
678 congestion statues, which are already available online for many cities worldwide. Mapping for  
679 spatial interpolation of particle concentrations for an urban area can visualise the pollution situation  
680 providing essential knowledge about air cleanliness, which is desired by residents, policymakers  
681 and built-environment professionals in order to secure the practical development of a healthy  
682 environment.

## 684 **Acknowledgement**

685 The research is supported by the China National Key R&D Programme SSHCool Project

686 ‘Solutions to Heating and Cooling of Buildings in the Yangtze River Region’ [Grant No.  
 687 2016YFC0700300] and the China Fundamental Research Funds for the Central Universities [Grant  
 688 No. 2018CDJDCH0015]. The research work is also based on the UK-China collaborative research  
 689 project ‘Low carbon climate-responsive Heating and Cooling of Cities (LoHCool)’ supported by the  
 690 National Natural Science Foundation of China [NSFC Grant No. 51561135002] and the UK  
 691 Engineering and Physical Sciences Research Council [EPSRC Grant No. EP/N009797/1]. The  
 692 authors would like to thank Prof. Howard Kipen and Dr Qingyu Meng from EOHSI, Rutgers  
 693 University for providing the technical guidance on the field measurement, Dr Han Wang, Ms  
 694 Tujingwa Zhang, Mr Zhu Chen and Mr Sheng Zhang for participating in the field measurement  
 695 campaign.

## 697 **References**

- 698 Ai, Z.T., Mak, C.M., 2013. CFD simulation of flow and dispersion around an isolated building: Effect of  
 699 inhomogeneous ABL and near-wall treatment. *Atmos. Environ.* 77, 568–578.  
 700 <https://doi.org/10.1016/J.ATMOSENV.2013.05.034>
- 701 Blocken, B., Janssen, W.D., van Hooff, T., 2012. CFD simulation for pedestrian wind comfort and wind  
 702 safety in urban areas: General decision framework and case study for the Eindhoven University campus. *Environ.*  
 703 *Model. Softw.* 30, 15–34. <https://doi.org/10.1016/j.envsoft.2011.11.009>
- 704 Chaloulakou, A., Grivas, G., Spyrellis, N., 2003. Neural network and multiple regression models for PM10  
 705 prediction in Athens: A comparative assessment. *J. Air Waste Manage. Assoc.* 53, 1183–1190.  
 706 <https://doi.org/https://doi.org/10.1080/10473289.2003.10466276>
- 707 China Meteorological Administration, n.d. Dataset of Daily Surface Observation Data in China [WWW  
 708 Document]. China Meteorol. Data Serv. Cent. URL  
 709 [http://data.cma.cn/data/cdcdetail/dataCode/SURF\\_CLI\\_CHN\\_MUL\\_DAY\\_V3.0.html](http://data.cma.cn/data/cdcdetail/dataCode/SURF_CLI_CHN_MUL_DAY_V3.0.html) (accessed 7.1.17).
- 710 China National Environmental Monitoring Centre, n.d. National Air Quality Real-time Release Platform  
 711 [WWW Document]. URL <http://106.37.208.233:20035/> (accessed 9.15.18).
- 712 Costanzo, V., Yao, R., Xu, T., Xiong, J., Zhang, Q., Li, B., 2019. Natural ventilation potential for residential  
 713 buildings in a densely built-up and highly polluted environment. A case study. *Renew. Energy* 138, 340–353.  
 714 <https://doi.org/10.1016/J.RENENE.2019.01.111>
- 715 de Gennaro, G., Trizio, L., Di Gilio, A., Pey, J., Pérez, N., Cusack, M., Alastuey, A., Querol, X., 2013.  
 716 Neural network model for the prediction of PM10 daily concentrations in two sites in the Western Mediterranean.  
 717 *Sci. Total Environ.* 463–464, 875–883. <https://doi.org/10.1016/j.scitotenv.2013.06.093>
- 718 Deligiorgi, D., Philippopoulos, K., 2011. Spatial Interpolation Methodologies in Urban Air Pollution  
 719 Modeling: Application for the Greater Area of Metropolitan Athens, Greece, in: *Advanced Air Pollution*. InTech,  
 720 Rijeka, pp. 341–362.

Department of Environment Food & Rural Affairs, n.d. UK AIR: Air Information Resource [WWW Document]. URL <https://uk-air.defra.gov.uk/> (accessed 9.15.18).

Dong, Y.H., Ng, S.T., 2015. A life cycle assessment model for evaluating the environmental impacts of building construction in Hong Kong. *Build. Environ.* 89, 183–191. <https://doi.org/10.1016/J.BUILDENV.2015.02.020>

Esri., 2019a. Spatial Join [WWW Document]. Spat. Join—Help | ArcGIS Deskt. URL <http://desktop.arcgis.com/en/arcmap/latest/tools/analysis-toolbox/spatial-join.htm> (accessed 3.15.19).

Esri., 2019b. Near [WWW Document]. Near—Help | ArcGIS Deskt. URL <http://desktop.arcgis.com/en/arcmap/latest/tools/analysis-toolbox/near.htm> (accessed 2.15.19).

Esri, 2018. What is Empirical Bayesian kriging? [WWW Document]. URL <http://pro.arcgis.com/en/pro-app/help/analysis/geostatistical-analyst/what-is-empirical-bayesian-kriging-.htm> (accessed 1.25.19).

European Environment Agency (EEA), 2018. Emissions of air pollutants from transport [WWW Document]. URL <https://www.eea.europa.eu/data-and-maps/indicators/transport-emissions-of-air-pollutants-8/transport-emissions-of-air-pollutants-6> (accessed 3.15.19).

Fan, Y. Van, Perry, S., Klemeš, J.J., Lee, C.T., 2018. A review on air emissions assessment: Transportation. *J. Clean. Prod.* 194, 673–684. <https://doi.org/10.1016/J.JCLEPRO.2018.05.151>

General Administration of Quality Supervision, Inspection and Quarantine, M. of, China, E.P. of, 2012. GB 3095-2012 Ambient air quality standards. China Environmental Science Press, Beijing.

Giovanis, E., 2018. The relationship between teleworking, traffic and air pollution. *Atmos. Pollut. Res.* 9, 1–14. <https://doi.org/10.1016/J.APR.2017.06.004>

Greater London Authority, 2014. The Control of Dust and Emissions During Construction and Demolition - Supplementary Planning Guidance. Greater London Authority, London.

Guilbert, A., De Cremer, K., Heene, B., Demoury, C., Aerts, R., Declerck, P., Brasseur, O., Van Nieuwenhuysse, A., 2019. Personal exposure to traffic-related air pollutants and relationships with respiratory symptoms and oxidative stress: A pilot cross-sectional study among urban green space workers. *Sci. Total Environ.* 649, 620–628. <https://doi.org/10.1016/J.SCITOTENV.2018.08.338>

Haykin, S.O., 2009. *Neural Networks and Learning Machines: A Comprehensive Foundation*, 3rd Editio. ed. Pearson Education.

He, H.-D., Lu, W.-Z., Xue, Y., 2015. Prediction of particulate matters at urban intersection by using multilayer perceptron model based on principal components. *Stoch. Environ. Res. Risk Assess.* 29, 2107–2114. <https://doi.org/10.1007/s00477-014-0989-x>

He, H., Lu, W.-Z., 2012. Urban aerosol particulates on Hong Kong roadsides: size distribution and concentration levels with time. *Stoch. Environ. Res. Risk Assess.* 26, 177–187. <https://doi.org/https://doi.org/10.1007/s00477-011-0465-9>

Health Effects Institute, 2010. *Traffic-Related Air Pollution: A Critical Review of the Literature on Emissions, Exposure, and Health Effects*. Boston.

Honarvar, A.R., Sami, A., 2019. Towards Sustainable Smart City by Particulate Matter Prediction Using Urban Big Data, Excluding Expensive Air Pollution Infrastructures. *Big Data Res.* 17, 56–65. <https://doi.org/10.1016/j.bdr.2018.05.006>

Ishak, A. Ben, Moslah, Z., Trabelsi, A., 2016. Analysis and prediction of PM10 concentration levels in Tunisia using statistical learning approaches. *Environ. Ecol. Stat.* 23, 469–490. <https://doi.org/https://doi.org/10.1007/s10651-016-0349-8>

Jacob, D.J., Winner, D.A., 2009. Effect of climate change on air quality. *Atmos. Environ.* 43, 51–63.  
<https://doi.org/10.1016/J.ATMOSENV.2008.09.051>

Jiang, F., 2013. Comparative study of the test results from aerosol monitoring method by DustTrak DRX and tapered element oscillating microbalance (TEOM).

Kim, K.-H., Jahan, S.A., Kabir, E., 2013. A review on human health perspective of air pollution with respect to allergies and asthma. *Environ. Int.* 59, 41–52. <https://doi.org/10.1016/J.ENVINT.2013.05.007>

Kim, M.J., Park, R.J., Kim, J.-J., Park, S.H., Chang, L.-S., Lee, D.-G., Choi, J.-Y., 2019. Computational fluid dynamics simulation of reactive fine particulate matter in a street canyon. *Atmos. Environ.* 209, 54–66.  
<https://doi.org/10.1016/j.atmosenv.2019.04.013>

Kuhn, M., 2018. CRAN - Package caret [WWW Document]. URL <https://cran.r-project.org/web/packages/caret/> (accessed 12.15.18).

Künzli, N., Kaiser, R., Medina, S., Studnicka, M., Chanel, O., Filliger, P., Herry, M., Horak, F., Puybonnieux-Texier, V., Qu  nel, P., Schneider, J., Seethaler, R., Vergnaud, J.-C., Sommer, H., 2000. Public-health impact of outdoor and traffic-related air pollution: a European assessment. *Lancet* 356, 795–801.  
[https://doi.org/10.1016/S0140-6736\(00\)02653-2](https://doi.org/10.1016/S0140-6736(00)02653-2)

Lateb, M., Meroney, R.N., Yataghene, M., Fellouah, H., Saleh, F., Boufadel, M.C., 2016. On the use of numerical modelling for near-field pollutant dispersion in urban environments – A review. *Environ. Pollut.* 208, 271–283. <https://doi.org/10.1016/J.ENVPOL.2015.07.039>

Lelieveld, J., Evans, J.S., Fnais, M., Giannadaki, D., Pozzer, A., 2015. The contribution of outdoor air pollution sources to premature mortality on a global scale. *Nature* 525, 367–371.  
<https://doi.org/10.1038/nature15371>

Li, B., Li, X.-B., Li, C., Zhu, Y., Peng, Z.-R., Wang, Z., Lu, S.-J., 2019. Impacts of wind fields on the distribution patterns of traffic emitted particles in urban residential areas. *Transp. Res. Part D Transp. Environ.* 68, 122–136. <https://doi.org/10.1016/j.trd.2018.01.030>

Li, X.-X., Liu, C.-H., Leung, D.Y.C., Lam, K.M., 2006. Recent progress in CFD modelling of wind field and pollutant transport in street canyons. *Atmos. Environ.* 40, 5640–5658.  
<https://doi.org/10.1016/J.ATMOSENV.2006.04.055>

Li, X., Huang, S., Jiao, A., Yang, X., Yun, J., Wang, Y., Xue, X., Chu, Y., Liu, F., Liu, Y., Ren, M., Chen, X., Li, N., Lu, Y., Mao, Z., Tian, L., Xiang, H., 2017. Association between ambient fine particulate matter and preterm birth or term low birth weight: An updated systematic review and meta-analysis. *Environ. Pollut.* 227, 596–605. <https://doi.org/10.1016/J.ENVPOL.2017.03.055>

Li, Z., Tang, Y., Song, X., Lazar, L., Li, Zhen, Zhao, J., 2019. Impact of ambient PM2.5 on adverse birth outcome and potential molecular mechanism. *Ecotoxicol. Environ. Saf.* 169, 248–254.  
<https://doi.org/10.1016/J.ECOENV.2018.10.109>

Mih  i  , A.S., Dupont, L., Chery, O., Camargo, M., Cai, C., 2019. Evaluating air quality by combining stationary, smart mobile pollution monitoring and data-driven modelling. *J. Clean. Prod.* 221, 398–418.  
<https://doi.org/10.1016/J.JCLEPRO.2019.02.179>

Nayebare, S.R., Aburizaiza, O.S., Siddique, A., Carpenter, D.O., Arden Pope, C., Mirza, H.M., Zeb, J., Aburiziza, A.J., Khwaja, H.A., 2019. Fine particles exposure and cardiopulmonary morbidity in Jeddah: A time-series analysis. *Sci. Total Environ.* 647, 1314–1322. <https://doi.org/10.1016/J.SCITOTENV.2018.08.094>

Oke, T.R., Mills, G., Christen, A., Voogt, J.A., 2017. *Urban Climates*. Cambridge University Press, Cambridge.



Özdemir, U., Taner, S., 2014. Impacts of Meteorological Factors on PM10: Artificial Neural Networks (ANN) and Multiple Linear Regression (MLR) Approaches. *Environ. Forensics* 15, 329–336. <https://doi.org/https://doi.org/10.1080/15275922.2014.950774>

Perez, P., Reyes, J., 2001. Prediction of Particulate Air Pollution using Neural Techniques. *Neural Comput. Appl.* 10, 165–171. <https://doi.org/10.1007/s005210170008>

Pope III, C.A., Burnett, R.T., Thun, M.J., Calle, E.E., Krewski, D., Ito, K., Thurston, G.D., 2002. Lung Cancer, Cardiopulmonary Mortality, and Long-term Exposure to Fine Particulate Air Pollution. *JAMA* 287, 1132–1141. <https://doi.org/10.1001/jama.287.9.1132>

R Core Team, 2018. An Introduction to R [WWW Document]. URL <https://cran.r-project.org/doc/manuals/r-release/R-intro.html> (accessed 12.15.18).

Ratti, C., Raydan, D., Steemers, K., 2003. Building form and environmental performance: archetypes, analysis and an arid climate. *Energy Build.* 35, 49–59.

Saeed, S., Hussain, L., Awan, I.A., Idris, A., 2017. Comparative Analysis of different Statistical Methods for Prediction of PM2.5 and PM10 Concentrations in Advance for Several Hours. *Int. J. Comput. Sci. Netw. Secur.* 17, 45–52.

Salim, S.M., Buccolieri, R., Chan, A., Di Sabatino, S., 2011. Numerical simulation of atmospheric pollutant dispersion in an urban street canyon: Comparison between RANS and LES. *J. Wind Eng. Ind. Aerodyn.* 99, 103–113. <https://doi.org/10.1016/J.JWEIA.2010.12.002>

Shi, K., Wang, H., Yang, Q., Wang, L., Sun, X., Li, Y., 2019. Exploring the relationships between urban forms and fine particulate (PM2.5) concentration in China: A multi-perspective study. *J. Clean. Prod.* 231, 990–1004. <https://doi.org/10.1016/J.JCLEPRO.2019.05.317>

Shieh, Y.-Y., Fouladi, R.T., 2003. The Effect of Multicollinearity on Multilevel Modeling Parameter Estimates and Standard Errors. *Educ. Psychol. Meas.* 63, 951–985. <https://doi.org/10.1177/0013164403258402>

Short, C.A., Song, J., Mottet, L., Chen, S., Wu, J., Ge, J., 2018. Challenges in the low-carbon adaptation of China’s apartment towers. *Build. Res. Inf.* 46, 899–930. <https://doi.org/10.1080/09613218.2018.1489465>

Stern, R., Builtjes, P., Schaap, M., Timmermans, R., Vautard, R., Hodzic, A., Memmesheimer, M., Feldmann, H., Renner, E., Wolke, R., Kerschbaumer, A., 2008. A model inter-comparison study focussing on episodes with elevated PM10 concentrations. *Atmos. Environ.* 42, 4567–4588. <https://doi.org/10.1016/j.atmosenv.2008.01.068>

Sun, C., Luo, Y., Li, J., 2018. Urban traffic infrastructure investment and air pollution: Evidence from the 83 cities in China. *J. Clean. Prod.* 172, 488–496. <https://doi.org/10.1016/J.JCLEPRO.2017.10.194>

Tai, A.P.K., Mickley, L.J., Jacob, D.J., 2010. Correlations between fine particulate matter (PM2.5) and meteorological variables in the United States: Implications for the sensitivity of PM2.5 to climate change. *Atmos. Environ.* 44, 3976–3984. <https://doi.org/10.1016/J.ATMOSENV.2010.06.060>

The Ministry of Housing and Urban-Rural Development of the People’s Republic of China, 2005. GB 50352-2005 Code for design of civil buildings. China Architecture & Building Press, Beijing.

Tian, J., Chen, D., 2010. A semi-empirical model for predicting hourly ground-level fine particulate matter (PM2.5) concentration in southern Ontario from satellite remote sensing and ground-based meteorological measurements. *Remote Sens. Environ.* 114, 221–229. <https://doi.org/10.1016/J.RSE.2009.09.011>

Tominaga, Y., Stathopoulos, T., 2011. CFD modeling of pollution dispersion in a street canyon: Comparison between LES and RANS. *J. Wind Eng. Ind. Aerodyn.* 99, 340–348. <https://doi.org/10.1016/j.jweia.2010.12.005>

Tong, Z., Chen, Y., Malkawi, A., Liu, Z., Freeman, R.B., 2016. Energy saving potential of natural

ventilation in China: The impact of ambient air pollution. *Appl. Energy* 179, 660–668.  
<https://doi.org/10.1016/J.APENERGY.2016.07.019>

United States Environmental Protection Agency, 2018. Health and Environmental Effects of Particulate Matter (PM) [WWW Document]. URL <https://www.epa.gov/pm-pollution/health-and-environmental-effects-particulate-matter-pm> (accessed 8.15.18).

Vicente, B., Rafael, S., Rodrigues, V., Relvas, H., Vilaça, M., Teixeira, J., Bandeira, J., Coelho, M., Borrego, C., 2018. Influence of different complexity levels of road traffic models on air quality modelling at street scale. *Air Qual. Atmos. Heal.* 11, 1217–1232. <https://doi.org/10.1007/s11869-018-0621-1>

Weinmayr, G., Pedersen, M., Stafoggia, M., Andersen, Z.J., Galassi, C., Munkenast, J., Jaensch, A., Oftedal, B., Krog, N.H., Aamodt, G., Pyko, A., Pershagen, G., Korek, M., De Faire, U., Pedersen, N.L., Östenson, C.-G., Rizzuto, D., Sørensen, M., Tjønneland, A., Bueno-de-Mesquita, B., Vermeulen, R., Eeftens, M., Concin, H., Lang, A., Wang, M., Tsai, M.-Y., Ricceri, F., Sacerdote, C., Ranzi, A., Cesaroni, G., Forastiere, F., de Hoogh, K., Beelen, R., Vineis, P., Kooter, I., Sokhi, R., Brunekreef, B., Hoek, G., Raaschou-Nielsen, O., Nagel, G., 2018. Particulate matter air pollution components and incidence of cancers of the stomach and the upper aerodigestive tract in the European Study of Cohorts of Air Pollution Effects (ESCAPE). *Environ. Int.* 120, 163–171. <https://doi.org/10.1016/J.ENVINT.2018.07.030>

World Health Organization, 2006. WHO Air quality guidelines for particulate matter, ozone, nitrogen dioxide and sulfur dioxide. WHO Press, Geneva.

Xu, X., Zhang, H., Chen, J., Li, Q., Wang, X., Wang, W., Zhang, Q., Xue, L., Ding, A., Mellouki, A., 2018. Six sources mainly contributing to the haze episodes and health risk assessment of PM<sub>2.5</sub> at Beijing suburb in winter 2016. *Ecotoxicol. Environ. Saf.* 166, 146–156. <https://doi.org/10.1016/J.ECOENV.2018.09.069>

Yao, R., Costanzo, V., Li, X., Zhang, Q., Li, B., 2018. The effect of passive measures on thermal comfort and energy conservation. A case study of the hot summer and cold winter climate in the Yangtze River region. *J. Build. Eng.* 15, 298–310. <https://doi.org/10.1016/j.jobbe.2017.11.012>

Yu, B., Liu, H., Wu, J., Hu, Y., Zhang, L., 2010. Automated derivation of urban building density information using airborne LiDAR data and object-based method. *Landsc. Urban Plan.* 98, 210–219. <https://doi.org/10.1016/j.landurbplan.2010.08.004>

Zhou, C., Li, S., Wang, S., 2018. Examining the impacts of urban form on air pollution in developing countries: A case study of China’s megacities. *Int. J. Environ. Res. Public Health* 15, 1–18. <https://doi.org/10.3390/ijerph15081565>

Zuo, J., Rameezdeen, R., Hagger, M., Zhou, Z., Ding, Z., 2017. Dust pollution control on construction sites: Awareness and self-responsibility of managers. *J. Clean. Prod.* 166, 312–320. <https://doi.org/10.1016/J.JCLEPRO.2017.08.027>